



(19) Europäisches Patentamt  
European Patent Office  
Office européen des brevets



(11) EP 0 948 237 A2

EUROPÄISCHE PATENTANMELDUNG

(51) Int. Cl. 6: H04R 3/00

(43) Veröffentlichungstag:  
06.10.1999 Patentblatt 1999/40

(21) Anmeldenummer: 99106123.5

(22) Anmeldetag: 01.04.1999

(84) Benannte Vertragsstaaten:  
AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU  
MC NL PT SE  
Benannte Erfindungsstaaten:  
AL LT LV MK RO SI

(71) Anmelder:  
DaimlerChrysler Aerospace AG  
81663 München (DE)  
(72) Erfinder: Thomas, Hans-Jörg  
89233 Neu-Ulm (DE)

(30) Priorität: 03.04.1998 DE 19814971

Szene Z4

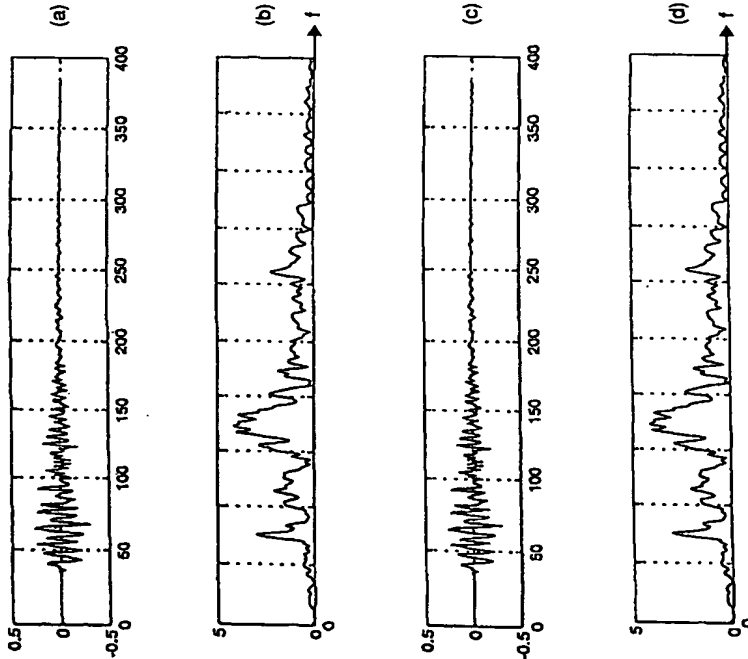


FIG.28

EP 0 948 237 A2

(54) Verfahren zur Störfreileitung eines Mikrophonesignals

(57) Für ein Verfahren zur Befreiung eines Mikrophonesignals von Störsignalen durch Erzeugen eines Kompensations Signals und Subtraktion des Kompensations Signals vom Mikrophonesignal wird vorgeschlagen, die Kompensation vollständig im Frequenzbereich vorzunehmen und auch das Ausgangssignal im Frequenzbereich zu verarbeiten. Es werden Maßnahmen zur Verringerung des Aufwands bei der Signalverarbeitung angegeben. Vorteile bei Weiterbildungen werden beispielsweise vor, daß zur

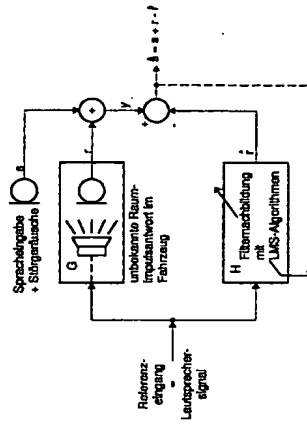


FIG.1

## Beschreibung

- [0001] Die Erfindung betrifft ein Verfahren zur Stöber-  
 [0002] treuung eines Mikrophonsignals.  
 [0003] Derartige Verfahren gewinnen insbesondere  
 [0004] für die Sprachübertragung von Kommandos und/oder für  
 [0005] Freisprechtelefone zunehmend an Bedeutung, wobei  
 [0006] insbesondere die Situation in einem Fahrzeug ein wich-  
 [0007] tiger Anwendungsfall ist.  
 [0008] Eine besondere Situation ist in Fahrzeugen  
 [0009] häufig dadurch gegeben, daß ein Widerlagereit wie  
 [0010] z.B. ein Radio, ein Kassetten- oder CD-Abspieler über  
 [0011] einen Lautsprecher eine Geräuschumgebung erzeugt,  
 [0012] die als Störquelle ein von einem Mikrophon aufgenom-  
 [0013] menes Sprachsignal beispielsweise für die Spracher-  
 [0014] kennung oder Telefonübertragung überlagert. Zur  
 [0015] Erkennung von Sprachanteilen in einem Sprachkanal  
 [0016] oder zur verständlichen Sprachübertragung über  
 [0017] Telefon ist das Mikrophonsignal soweit wie möglich von  
 [0018] Störanteilen zu befreien.  
 [0019] Das von einer Störquelle, insbesondere einem  
 [0020] Lautsprecher ausgehende Störsignal gelangt nicht nur  
 [0021] auf direktem kürzesten Weg zum Mikrophon, sondern  
 [0022] tritt auch noch über zahlreiche Reflexionen als eine  
 [0023] Überlagerung einer Vielzahl von Echos mit verschie-  
 [0024] denen Laufzeiten im Mikrophonsignal in Erscheinung.  
 [0025] Die gesamte Einwirkung des Störsignals von der Stör-  
 [0026] quelle auf das Mikrophonsignal kann durch eine a priori  
 [0027] unbekannte Übertragungsfunktion des Raumes, be-  
 [0028] speziell des Fahrzeugaumes eines Kraftfahrzeugs  
 [0029] beschrieben werden. Die Übertragungsfunktion ändert  
 [0030] sich je nach Besetzung des Fahrzeugs und nach Posi-  
 [0031] tion der einzelnen Personen. Durch Nachbildung dieser  
 [0032] Übertragungsfunktion und Filterung eines Referenzsi-  
 [0033] gnals von der Störquelle mit dieser Nachbildung kann  
 [0034] ein Kompensationsignal erzeugt werden, welches  
 [0035] durch Subtraktion vom Mikrophonsignal ein vom Stör-  
 [0036] signal bereinigtes Signal beispielsweise ein reines Sprach-  
 [0037] signal liefert. Im Realfall stellt die genannte  
 [0038] Nachbildung eine mehr oder minder gute Annäherung  
 [0039] an die unbekannte Übertragungsfunktion dar und die  
 [0040] Störung kann nicht vollständig beseitigt werden.  
 [0041] Aufgabe der vorliegenden Erfindung ist es, ein  
 [0042] Verfahren zur Stöberreinigung eines Mikrophonsignals  
 [0043] anzugeben, daß bei vertretbarem Signalverarbeitungs-  
 [0044] aufwand gute Eigenschaften hinsichtlich der Erfindung  
 [0045] aufweist.  
 [0046] Die Erfindung ist im Patentanspruch 1  
 [0047] beschrieben. Die Unteransprüche enthalten vorteilhafte  
 [0048] Ausgestaltungen und Weiterbildungen der Erfindung.  
 [0049] Wesentlich an dem erfindungsgegenständlichen Ver-  
 [0050] fahren ist, daß die Kompensation des Störanteils  
 [0051] im Mikrophonsignal mittels eines aus dem Referenzsi-  
 [0052] gnal über die Nachbildung der Übertragungsfunktion  
 [0053] erzeugten Kompensationssignals im Frequenzbereich  
 [0054] vorgenommen wird, so daß Mikrophonsignal, Kompen-  
 [0055] sationsignal und Ausgangssignal im Frequenzbereich,  
 [0056] d.h. in Form von Spektren vorliegen. Die Signalverar-  
 [0057] beitung in diesem Verfahrensschritt im Frequenzbereich  
 [0058] erfordert zwar eine spätere Transformation des Mikro-  
 [0059] phonsignals, beispielsweise aber, daß die Nachbildung  
 [0060] der Übertragungsfunktion im Frequenzbereich vorteil-  
 [0061] hafter ist und stellt für eine vorteilhafte nachfolgende  
 [0062] zusätzliche Gerätschreduktion des Ausgangssignals,  
 [0063] die typischerweise ebenfalls im Frequenzbereich vor-  
 [0064] genommen wird, bereits eine besonders geeignete  
 [0065] Signalform bereit.  
 [0066] Durch einfache Näherungen beim Ersatz eines  
 [0067] Verarbeitungsschritts mit einem Zeitfilter kann durch  
 [0068] Übergang zu einer Fühung im Frequenzbereich eine  
 [0069] deutliche Reduzierung des Verarbeitungsaufwands  
 [0070] reduziert werden.  
 [0071] Für lange Impulsantworten der Übertrags-  
 [0072] funktion bzw. deren Nachbildung sieht eine vorteilhafte  
 [0073] Weiterbildung der Erfindung eine Aufteilung des Nach-  
 [0074] bildungssignals in mehrere Teilliter zu zeitversetzten  
 [0075] Segmenten des segmentierten Referenzsignals vor,  
 [0076] deren Koeffizienten-Aktualisierung zeitlich gestaffelt  
 [0077] sein kann, wodurch der Signalverarbeitungsaufwand  
 [0078] gering gehalten werden kann.  
 [0079] Als besonders vorteilhaft erweist es sich, die  
 [0080] Erfindung eines Sprachsignals auf der Basis einer Ein-  
 [0081] stellung des Nachbildungssignals, die in einer vorherge-  
 [0082] henden Sprachpause gewonnen und gespeichert  
 [0083] wurde, vorzunehmen.  
 [0084] Die Aufteilung des Nachbildungssignals in meh-  
 [0085] rere Teilliter und die Stöberreinigung auf der Basis einer  
 [0086] in einer Sprachpause gewonnenen Filtereinstellung  
 [0087] sind auch unabhängig von der Störquellenkompensation  
 [0088] im Frequenzbereich eigenständig für die Stöberreinigung  
 [0089] eines Mikrophonsignals realisierbar und vorteilhaft.  
 [0090] Die Erfindung ist nachfolgend anhand von  
 [0091] bevorzugten Ausführungsbeispielen unter Bezugnahme  
 [0092] auf die Abbildungen noch eingehend veranschaulicht.  
 [0093] Dabei zeigt:

- Fig. 1 ein Prinzip der Kompensation eines Radio-  
 [0094] signals  
 Fig. 2a ein Blockschaltbild zu Fig. 1  
 Fig. 2b ein Blockschaltbild zur Filternachbildung  
 Fig. 3 ein detailliertes Beispiel zu Fig. 2b  
 Fig. 4 eine Erweiterung auf mehrere Teilliter  
 Fig. 5 einen Übergang zur Kompensation im Fre-  
 [0095] quenzbereich  
 Fig. 6 ein detailliertes Beispiel zu Fig. 5b  
 Fig. 7 ein Ausführungsbeispiel mit mehreren Teil-  
 [0096] litern  
 Fig. 8 ein Ausführungsbeispiel mit Speicherung

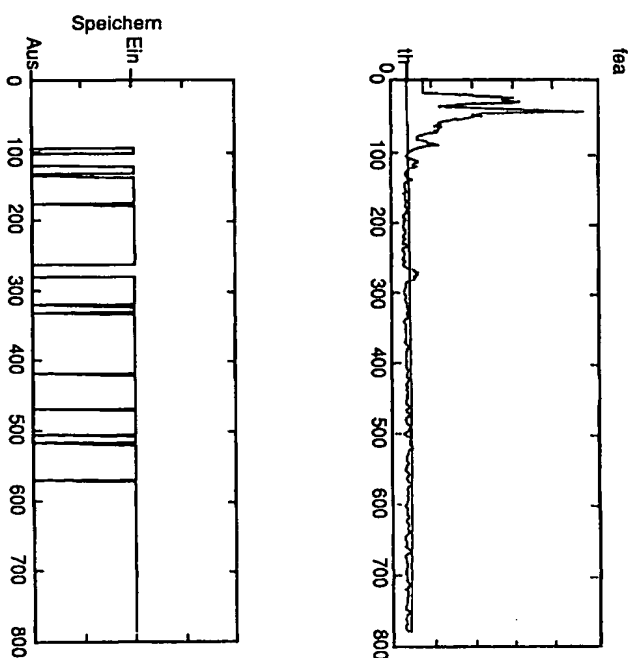


FIG.27

Szene Z4

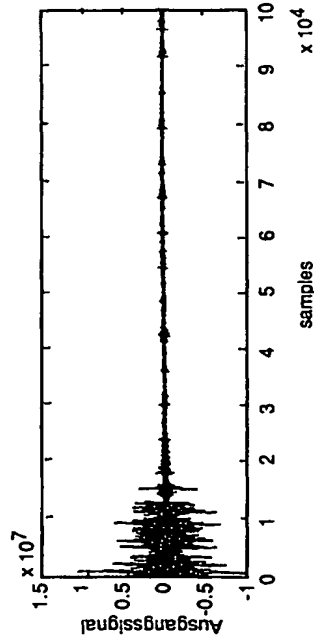
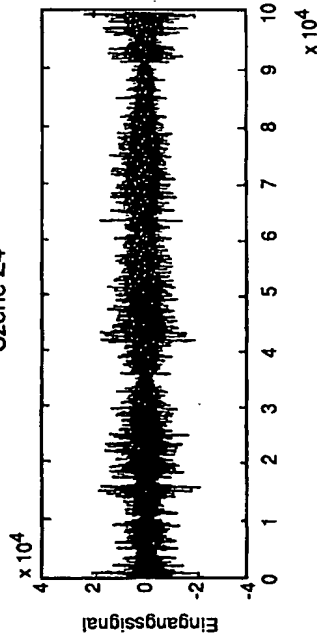


FIG.26

der Filtereinstellungen

Fig. 9 Signale einer synthetischen Beispielszene

Fig. 10 Impulsantwort und Übertragungsfunktion zu Fig. 9

Fig. 11 Signal einer ersten Meßszene

Fig. 12 Impulsantwort und Übertragungsfunktion zu Fig. 11

Fig. 13 das Beispiel nach Fig. 11 mit Speicherung der Filtereinstellungen

Fig. 14 eine Sprachpausendetektion zu Fig. 13

Fig. 15 Impulsantworten und Übertragungsfunktionen zu Fig. 11 und Fig. 13

Fig. 16 Übergang von einem Zeitfenster zu einer Faltung im Frequenzbereich

Fig. 17 ein Rechteck-Zeitfenster mit UnTERSPEKtrum

Fig. 18 ein Hamming-Zeitfenster mit UnTERSPEKtrum

Fig. 19 Staffeung von Signalblöcken bei der Filterberechnung

Fig. 20 Signale einer zweiten Meßszene

Fig. 21 eine Sprachpausendetektion zu Fig. 20

Fig. 22 Impulsantworten und Übertragungsfunktionen zu Fig. 20 und Fig. 21

Fig. 23 Signale einer dritten Meßszene

Fig. 24 eine Sprachpausendetektion zu Fig. 23

Fig. 25 Impulsantworten und Übertragungsfunktionen zu Fig. 23 und Fig. 24

Fig. 26 Signale einer vierten Meßszene

Fig. 27 eine Sprachpausendetektion zu Fig. 26

Fig. 28 Impulsantworten und Übertragungsfunktionen zu Fig. 26 und Fig. 27.

[0013] Fig. 1 stellt das Prinzip einer Einrichtung zur (einkanaligen) Radiosignalkompensation dar. Das vom Lautsprecher abgestrahlte akustische Signal gelangt auf direktem Wege, aber auch über zahlreiche Reflexionen im Fahrzeuginnenraum, auf das Mikrofon des

Sprachengabesystems. Unter der Annahme, daß sich die Übertragungsstrecke G demnach als Transversalfilter mit einer gewichteten Summe zeitlich verzögerter Echos darstellt, läßt sich eine Filternachbildung H finden, die im Idealfall  $H=1/G$  eine vollständige Kompensation des Radiosignales ermöglicht.

[0014] Das Lautsprecher-signal  $x$  wird durch die a priori unbekannte Übertragungsfunktion G des Fahrzeuginnenraumes geleitet. Es entsteht die Störkomponente  $r$ , die sich mit dem Sprachsignal  $s$  zu dem Mikrophonsignal  $y$  addiert. Um die Störkomponente  $r$  zu kompensieren, wird mittels der Filternachbildung H ein Schätzwert  $r^*$  aus dem Lautsprecher-signal  $x$  erzeugt. Der Ausgang der Schaltung liefert den Schätzwert für das Sprachsignal:

$$\hat{s}^* = s + r - r^* = s - E$$

[0015] Dem Sprachsignal  $s$  ist also am Ausgang der Schaltung noch das Fehlersignal  $E = r - r^*$  überlagert, welches in der Praxis möglichst klein gehalten werden sollte. Das Sprachsignal kann noch Störungen in Form von z.B. Motorgerauschen oder externen Geräuschen enthalten, die aber in diesem Zusammenhang nicht explizit behandelt werden.

[0016] H ist ein adaptives Filter und arbeitet nach einem in der Literatur bekannten Standardverfahren, dem LMS-Algorithmus (least mean squares). Neben dem Eingangssignal  $x$  wird noch das Fehlersignal E benötigt, um die Koeffizientenadaptation im Filter H zu bewerkstelligen. Hierfür ist das Ausgangssignal  $s^*$  der Bestimmung der Filterkoeffizienten zugeführt.

[0017] Fig. 2a zeigt in anderer Darstellung nochmals die Anordnung von Fig. 1 als Radiosignalkompensation. Das adaptive System H kann z.B. im Zeitbereich als FIR-Filter (finite-impulse-response-Filter) realisiert werden. Bei großen Impulsantwortlängen, wie sie in der Praxis häufig auftreten, ist hierzu allerdings ein sehr hoher Rechenaufwand notwendig. Verschiedene Vor- teile gegenüber einer Zeitbereichslösung bietet die Realisierung des LMS-Algorithmus im Frequenzbereich (FLMS). Wegen der blockweisen Verarbeitung von Daten in den als diskreten Fouriertransformationen realisierten spektralen Transformationen und der Filterrealisierung im Frequenzbereich durch Multiplikationen wird dieses Verfahren besonders recheneffizient.

[0018] Fig. 2b zeigt ein Blockschaltbild des FLMS-Algorithmus. Die zugehörige Theorie ist an sich bekannt und daher an dieser Stelle nicht im Detail behandelt. Es bedeuten F eine spektrale Transformation FFT eines Zeitsignals in den Frequenzbereich und  $F^{-1}$  die inverse IFFT. Die als Projektionen P1, P2 und P3 bezeichneten Verarbeitungsschritte dienen der korrekten Segmentierung der Daten durch die blockweise Verwendung mit der FFT bzw. IFFT und werden später noch genauer erläutert. Die Abteilweise des Filters besteht in der Multiplikation des Referenzspektrums X mit dem Filter-Koeffizientenvektor H. Das Spektrum des

Filterausgangs  $P^*$  wird über  $P^*$  zurück in den Zeitbereich transformiert. Nach Anwendung der Projektion  $P_2$  auf den Realteil des so erhaltenen Kompensations Signals steht das Signal  $r^*$  zur Verfügung. Die Differenz der Signale

$$s^* = y^* - r^* = s + r - r^* = s + E$$

stellt den eigentlichen Ausgang, eine Schätzung der Sprachengabe, dar.

[0019] Wesentlicher Bestandteil des adaptiven Filters ist die Koeffizientenadaptation im Block K, die im Fig. 2b durch die Erneuerungsgleichung

$$H = H' + \Delta H'$$

beschrieben wird. Die hier mit zwei spektralen Tensoroperationen besonders aufwendige Projektion  $P_1$  berechnet aus  $H$  den für die Filterung benötigten Koeffizientenvektor  $H'$ . Zur Berechnung des Korrekturvektors  $\Delta H'$  wird neben dem Referenzspektrum  $X$  das Spektrum  $S^*$  des mit  $P_2$  bewerteten Ausgangssignals  $s + r - r^*$  benötigt.

[0020] Ein detailliertes Blockschaltbild des in Fig. 2b dargestellten FLMS-Algorithmus zeigt Fig. 3. Die Abtastwerte eines Signals und die Stichstellen der FFT sehen in geträufelter Weise als samples bezeichnet. Alle Spektralinformationen und deren Inverse sind als 256-Punkte - FFTs, die jeweils um 128 samples überlappen, zu segmentieren zu beachten ist, daß sich das Ausgangssignal  $s^*$  im Zeitbereich aus 128-sample-Blocken zusammensetzt. Es entsteht aus der Differenz der zweiten Blockhälften (also jeweils der samples 129 bis 256) von Mikrosignal und gefiltertem Kompensationsignal  $r^*$ . Aufwendig ist die Projektion  $P_1$ , die 2 FFTs benötigt und den Vektor  $H'$  in den Vektor  $H$  umrechnet. Hierbei wird aus dem komplexen 256-Punkte-Eigenvektor der Rücktransformation vom Frequenz in den Zeitbereich (IFFT), die erste Hälfte (samples 1 bis 128) ausgeschnitten und die zweite Hälfte (samples 129 bis 256) zu Null gesetzt. Nach Anwendung dieses Rechteckentzerrers im Zeitbereich erfolgt wieder mittels FFT die Transformation in den Frequenzbereich. Einfach ist die Projektion  $P_2$ . Sie besteht aus der oben schon beschriebenen Ausschnittbildung der letzten 128 samples, wodurch aus überlappenden 256-sample-Blocken wieder nicht überlappende 128-sample-Blocke entstehen. Ebenfalls sehr einfach ist schließlich auch die Projektion  $P_3$ , welche umgekehrt aus nicht überlappenden 128-sample-Blocken des Ausgangssignals durch Verschieben von 128 Nullwerten wiederum überlappende 256-sample-Blocke bereitstellt. Die Adaption der Filterkoeffizienten  $H_{L+1}$  für einen Zyklus  $L+1$  besteht aus der Addition eines Erneuerungsvektors  $\Delta H_L$  zum alten Koeffizientenvektor  $H_L$ . Diese Erneuerung errechnet sich aus dem Produkt zwischen dem Spektrum  $S^*$  des Ausgangssignals und dem konjugiert komplexen Spektrum  $X_L^*$  des Referenz-

signals, gewichtet mit einer spektralen Leistungsnormierung  $2H_L$ ,  $\Delta H_L = 2H_L \cdot X_L^* \cdot S^*$ . Zum Zweck dieser Leistungsnormierung ist der mit einer Konstanten  $Z_0$  multiplizierte Kehrwert des quadrierten Referenzleistungsspektrums  $S_{X,L}$  zu berechnen  $2H_L = Z_0/S_{X,L}$ , wozu ein rekursives Filter 1. Ordnung mit einer Konstanten  $\beta$  dient

$$S_{X,L} = \beta \cdot |X_L|^2 + (1-\beta) \cdot S_{X,L-1}$$

[0021] Die Arbeitsweise des LMS-Algorithmus wird einheitlich von der Adaptionskonstante  $\alpha$  und der Gattungskonstante  $\beta$  beeinflusst. Zwischenspeicher in Rekursionsschritten sind mit  $S_0$  bezeichnet.

[0022] Die bisher beschriebene Anordnung des FLMS-Algorithmus erlaubt Filteranordnungen mit einer maximalen Impulsantwortlänge von einer halben FFT-Länge, im Beispielstil also 128 samples. Solche längere Impulsantworten kompensiert werden, ist der schon bekannte FLMS-Algorithmus für einen Teillfilter (Fig. 4a) auf  $n$  Teillfilter zu erweitern. Eine 3-Teillfilter-Lösung mit einer Impulsantwortlänge von  $3 \cdot 128 = 384$  samples hat sich bei der Radiosignalunterdrückung im Pkw mit einem Sprachengabesystem bewährt (Fig. 4b). Der im Fig. 4a mit B bezeichnete Block mit den Eingangssignalen  $X$  und  $S^*$  und dem Kompensations-Spektrum  $H'$  als Ausgang ist durch die im Fig. 4b dargestellte Erweiterung zu ersetzen. Das Spektrum  $X$  des Referenzsignals wird durch Zwischenspeicher  $D$  um 1 bzw. 2 Blocklängen verzögert und das unverzögerte  $X_1$  und die beiden verzögerten Speichen  $X_2$ ,  $X_3$  werden separat in mit in einer erweiterten Projektion  $P_1$  getrennt bestimmten Koeffizientenvektoren  $H_1$ ,  $H_2$ ,  $H_3$  multipliziert. Die Bildung der Koeffizientenvektoren erfolgt analog zum Fall nur eines Teillfilters, wobei in  $K_1$ ,  $K_2$ ,  $K_3$  jeweils das zugehörige Referenzspektrum mit dem Spektrum  $S^*$  des Ausgangssignals verknüpft wird. Der Aufwand wird hauptsächlich durch die Verdichtung der Projektion  $P_1$  beträchtlich erhöht. Zusätzlich Speicherplatzbedarf wird notwendig um die Speichen des um 1 bzw. 2 Blocklängen zeitlich älteren Referenzsignals  $X$  bereitzustellen.

[0023] Bei der beispielhaft angegebenen Aufgabenstellung der Unterdrückung des Radiosignals bei Sprachengabe im Kfz ist es vorteilhaft die Ausgangssigna nicht im Zeit-, sondern im Frequenzbereich auszugeben, da dadurch eine verbesserte Anpassung an eine nachgeschaltete Gerätsunterdrückung erreicht werden kann. Der bereits vorgestellte FLMS-Algorithmus mit einem Teillfilter benötigt gemäß Fig. 5a insgesamt 5 FFTs bei einem Ausgangssignal im Zeitbereich. Wird dem Ausgang eine FFT nachgeschaltet, erhöht sich der Aufwand bei einem Frequenzbereichs-Ausgangssignal auf 6 FFTs. Die gleiche FFT-Anzahl ergibt sich zunächst auch bei einer äquivalenten Lösung nach Fig. 5b. Diese Variante besitzt jedoch folgende Vorteile:

- Bei der zeitgleichen Spektralanalyse der Signale  $x$

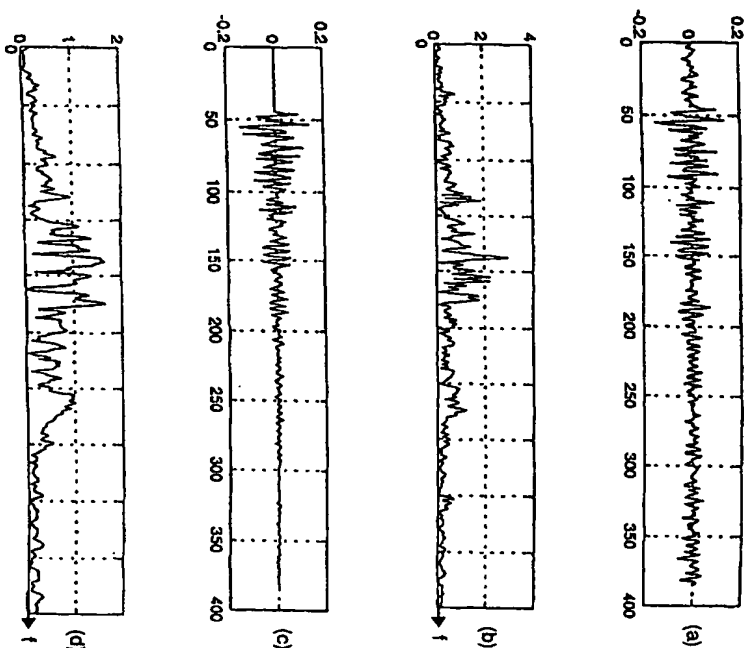


FIG.25



und Umrechnung in den Zeitbereich ergibt sich das in Fig. 11 unten dargestellte Ausgangssignal. Der Hörttest ergibt eine deutliche Heraushebung des Sprachanteils bzw. eine vor allem in den kurzen Sprachpausen bemerkenswerte Musikanterdrückung. Auffällig und von Nachteil ist jedoch, daß die erwünschte Radiosignallunterdrückung in starken Maße davon abhängt, ob gerade gesprochen wird oder nicht. Die wieder am Sender ermittelte 384-sample-Impulsantwort mit zugehöriger Übertragungsfunktion ist aus Fig. 12 ersichtlich. Eine korrekte Impulsantwort ist an den typischen Nullsamples (Töcke) am Anfang zu erkennen, welche von der Laufzeit des Direktschalls vom Radiolautsprecher zum Mikrofon herrühren. Aus den hier vorhandenen starken Störungen am Anfang sowie am Ende der Impulsantwort läßt sich demnach der Schluß ziehen, daß die Filteradaptation an dieser Stelle wegen vorhandener Spracheingabe äußerst unzureichend ist.

[0030] Die im folgenden anhand von Fig. 8 beschriebene Ausführungsform beruht auf folgender Grundidee: Unterbreitet das Merkmal für eine Spracheingabe, ein geeignetes Merkmal dient zusammen mit einem Schwellenwert als Indikator für eine Sprachpause. Unterbreitet das Merkmal die Schwelle, so ist dies ein Anzeichen für fehlende Spracheingabe. In diesem Fall kann - wie oben schon festgestellt - eine weitgehend ungestörte Filteradaptation erfolgen. Bei Spracheingabe wird nun auf denjenigen Filterkoeffizientensatz zurückgegriffen, der unmittelbar vor der Schwellenüberschreitung - d.h. am Ende der vorangegangenen Sprachpause - abgespeichert wurde. Diese gespeicherten Koeffizienten H10, H20, H30 liefern im Regelfall eine deutlich bessere Radiosignal-Kompensation als die unter dem störenden Einfluß der Spracheingabe sich ständig ändernden aktuellen Koeffizienten H, H2, H3.

[0031] Fig. 8 stellt eine Ausführungsform mit einer weiter verbesserten FLMMS-Verarbeitung mit 3 Teilfiltern dar. Neben den schon in Fig. 7 vorhandenen aktuellen Filterkoeffizientenvektoren H1, H2, H3, welche zur Bildung des fortlaufend adaptierten Ausgangssignals y-R benützt wurden, existiert nun ein zusätzliches Ausgangssignal (y-R0), das unter Verwendung gespeicherter Koeffizienten H10, H20, H30 gebildet ist. Die aktuellen Koeffizientensätze H1, H2, H3 stellen nur bei fehlender Spracheingabe im eingeschwungenen Zustand ein brauchbares Kompensationsfilter im Frequenzbereich dar, liefern hingegen bei Spracheingabe ungenügende Filtereigenschaften, weil der Adaptionsprozeß in der Regelschleife ständig gestört wird. Bei fehlender Spracheingabe d.h. hoher Filterqualität sind die drei Schalter geschlossen und es werden die aktuellen Koeffizientensätze in die Koeffizientenspeicher M1, M2, M3 geschrieben: H10=H1, H20=H2, H30=H3. Die Ausgangs (y-R0) und (y-R1) sind identisch. Einsetzende Sprachereignisse bewirken ein Öffnen der 3 Schalter, wodurch die zuletzt in den Speichern M1, M2, M3 befindlichen Koeffizienten H10, H20, H30 nicht mehr überschrieben werden und unverändert bleiben. Dieser

Zustand, in welchem sich die Ausgänge (y-R0) und (y-R1) unterscheiden, wird solange beibehalten, bis wieder eine Sprachpause detektiert und die Schalter geschlossen werden.

[0032] Als Sprachpausenmerkmal (ea) hat sich die geglättete Summe aller Absolutwerte der Koeffizientenkorrekturvektoren AH1', AH2', AH3 bewährt (Fig. 8a). Diese Größe ist gleich Null bzw. weist kleine Zahlenwerte auf, wenn es keinen oder nur einen geringen Bedarf gibt, die Koeffizienten abzuändern. In Sprachpausen ist dies der Fall, der Regelkreis ist praktisch eingeschwungen. Störungen, wie sie durch Sprachereignisse - aber auch durch Bewegungen der Fahrzeuginsassen - hervorgerufen werden, haben einen erhöhten Nachregelbedarf zur Folge, was sich durch entsprechend große Zahlenwerte bei AH1', AH2', AH3 und somit beim Merkmal (ea) bemerkbar macht. Ein Glimmfilterschaltkreis bewirkt eine rekursive Tiefpaß 1. Ordnung mit dem Eingang (ea) stellt an seinem Ausgang das geglättete Sprachpausen-Merkmal (ea) zur Verfügung, welches nach Vergleich mit einem Schwellenwert in die Schalter für die Koeffizientenübernahme steuert.

[0033] Die Wirkungsweise des verbesserten FLMMS-Algorithmus nach Fig. 8 demonstriert Fig. 13. Oben ist das aufgezeichnete Signal y der Szene Z1 (vgl. Fig. 11 oben) dargestellt, unten das gewonnenen Ausgangssignal. Schon der visuelle Vergleich der Ausgangssignale von Fig. 13 und Fig. 11 zeigt die verbesserte Heraushebung der Sprachpassagen. Der verglichende Hörtest bestätigt dies; auch während der Sprachereignisse ist die Musikanterdrückung deutlich besser. Den Verlauf des Sprachpausenmerkmals und der konstanten Schwelle über der Zeit (hier in FFT-Blick) stellen zeigt Fig. 14 oben. In den durch die Schwellenunterschreitung detektierten Sprachpausen (Fig. 14 unten) findet laufend die Übernahme der Koeffizienten in die Speicher wie beschrieben statt, um dort während der Sprachereignisse als gespeicherte Koeffizienten zur Verfügung zu stehen. Die schon in Fig. 12 am Szenenende gemessene 384-sample-Impulsantwort mit zugehöriger Betragsübertragungsfunktion ist in Fig. 15 als aktuelle Impulsantwort (a) bzw. aktuelle Übertragungsfunktion (b) dargestellt. Im Gegensatz zu dieser mögliche Sprachereignisse stark gestörten Schätzung aus den aktuellen Koeffizienten H1, H2, H3 ist aus den gespeicherten Koeffizienten H10, H20, H30 eine Impulsantwort (c) und eine Übertragungsfunktion (d) hoher Qualität berechenbar. Die Impulsantwort aus den gespeicherten Koeffizienten weist die typischen Nullsamples am Anfang auf, welche durch die Laufzeit des Direktschalls vom Radiolautsprecher zum Spracheingabemikrofon verursacht werden. Aus der im Beispiel abtastenden Taktzeit von ca. 40 samples läßt sich die Entfernung zwischen Lautsprecher und Mikrofon bestimmen.

[0034] Wie vorstehend schon angedeutet läßt sich die aufwendige Projektion P4 (FFT, Fenster, rechts im Zeitbereich, FFT) ohne merkliche Einbuße an Qualität

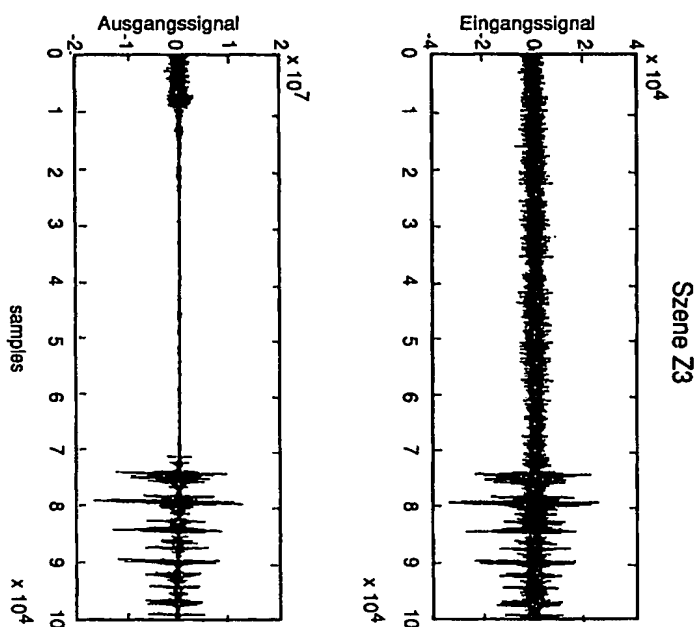


FIG.23

## Szene Z2

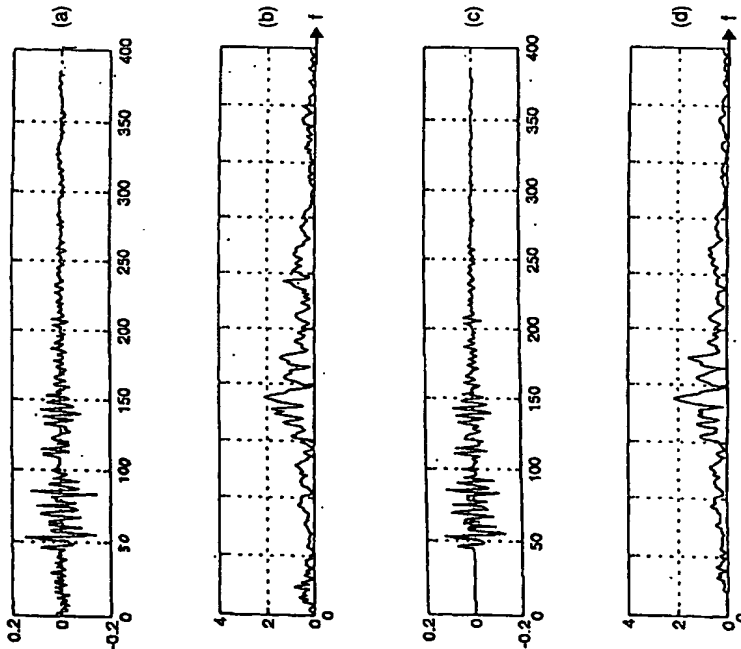


FIG.22

durch eine relativ einfache Faltung im Frequenzbereich ersetzt, wodurch 2 FFT's eingespart werden. Dazu beachte man Fig. 16. In einem ersten Schritt wird das "rechtsseitige" 128-sample-Rechteckfenster im Zeitbereich (Fig. 16a) bei der idealen Projektion ersetzt durch ein 128-sample-Hammingfenster (Fig. 16b). Gegenüber dem Rechteckfenster besitzt dieses den Vorteil eines bedeutend schmaleren Spektrums. Wie Fig. 17 zeigt, besteht beim Rechteckfenster der Realteil des Spektrums aus einer einzigen Linie (Gleichanteil), während das zur Mitte antisymmetrische Imaginärteil-Spektrum aus vielen nach außen hin langsam abfallenden Linien mit alternierenden Nullstellen besteht. Im Gegensatz dazu beschränkt sich das komplexe Spektrum des Hammingfensters (Fig. 18) auf insgesamt 7 Linien, von denen im symmetrischen Realteil nur 3 und im antisymmetrischen Imaginärteil nur 4 Werte von Null verschieden sind. Sämtliche weiter außen liegenden Anteile sind vernachlässigbar gering. Diese spezielle Eigenschaft des Hammingfensters ermöglicht es vorteilhaftweise die Multiplikation im Zeitbereich (Fig. 16b) zu ersetzen durch eine Faltung mit dem zugehörigen 7-sample-Spektrum im Frequenzbereich und damit eine IFFT und eine FFT einzusparen (Fig. 16c).

[0035] Prinzipiell läßt sich natürlich auch die Projektion P1 (IFFT - linksseitiges Rechteckfenster - FFT) ersetzen durch eine entsprechende Faltungsoperation im Frequenzbereich mit dem konjugiert komplexen 7-Linien-Spektrum. Experimente haben jedoch gezeigt, daß Einsparungen an dieser Stelle erkaufte werden mit einer deutlichen Verschlechterung des Einschwingverhaltens. Aufwandsreiche Lösungen lassen sich trotzdem dadurch erzielen, daß in dem LMS-Algorithmus nach Fig. 8 die 3 Projektionen P1 nicht gleichzeitig in einem 256-sample-Inputdatenblock abgearbeitet werden müssen. Die mit 128-samples überlappenden Inputdatenblöcke der Länge 256 sind mit einer willkürlich bei "1" beginnenden Nummerierung in Fig. 19a skizziert. So ist es z.B. möglich bei modulo-3-Zählweise der Inputdatenblöcke die 3 Teilfilterprojektionen nicht parallel (Fig. 19b) sondern sequentiell in aufeinanderfolgenden Blöcken Fig. 19 zu berechnen. Dadurch sind bei idealer Projektion P1 pro Datenblock nicht 6 sondern nur noch 2 FFT's notwendig. Es hat sich gezeigt, daß die Kompensation des Radiosignales auch noch ausreichend funktioniert, wenn die Abstände zwischen den zu berechnenden Teilfilterprojektionen noch größer gewählt werden. Zählt man die Blöcke z.B. modulo 6, so ist lediglich in jedem zweiten Block eine Projektion zu berechnen (Fig. 19c). Selbst eine Reduzierung auf einen Abstand von vier Blöcken zwischen zwei aufeinanderfolgenden P1 Berechnungen mittels modulo-12-Zählung führt noch zu brauchbaren Ergebnissen (Fig. 19d).

[0036] Die Leistungsfähigkeit des FLMS-Algorithmus mit 3 Teilfiltern gemäß Blockschaltung Fig. 8 und einer sequentiellen Berechnung der idealen Projektion P1 im Zeitalter nach Fig. 19e sowie der Projektion P2 mittels

Faltung im Frequenzbereich (Fig. 16c) mit einem komplexen 7-Linien-Spektrum (Fig. 18) sei anhand von 3 Meßbeispielen demonstriert.

[0037] Die erste dieser Szenen Z2 beinhaltet Sprach-eingabe von Ziffern, wobei der Radiosprecher annähernd weißes Rauschen mit verhältnismäßig hoher Lautstärke abstrahlt. Das zugehörige 10000-sample-Mikrophonsignal ist in Fig. 20 oben, das extrahierte Ausgangssignal in Fig. 20 unten dargestellt. Eine deutliche Sprachbereinigung des Outputsignales gegenüber dem Mikrophoneing stellt man durch Abhörvergleich fest. Der zeitliche Verlauf des Sprachpausenmerkmals ist zusammen mit der konstanten Schwelle in Fig. 21 oben abgebildet und die hieraus abgeleiteten Sprachpausen bzw. die zugeordneten Schalterstellungen in Fig. 21 unten. Schließlich zeigt Fig. 22 in zu Fig. 15 analoger Weise die am Szenenende gefundene Impulsantwort (a) und Übertragungsfunktion (b) auf der Basis der aktuellen Koeffizienten und die entsprechenden Größen (c), (d) auf der Basis der Sprachpauseneinstellung. Es ist deutlich erkennbar, daß die am Szenenende gefundene aktuelle Impulsantwort ein intigres Sprach-eingabe gestörtes Ergebnis darstellt, während die aus der letzten Sprachpause stammende Impulsantwort aus den gespeicherten Koeffizientensätzen eine hohe Qualität aufweist.

[0038] Die ersten 10000 samples einer Meßszene Z3 mit POP-Musik im Radio und flüssig bis schnell gesprochener Sprache der rechts hinten sitzenden Personen sind in Form des Mikrophonsignales  $y$  in Fig. 23 oben aufgezeichnet. Nach ca. 10000 samples (0.83 s) wird das Radiosignal brauchbar unterdrückt (Fig. 23 unten). Auch bei der im letzten Drittel dieser Szene einsetzenden Sprach-eingabe bleibt die POP-Musikunterdrückung wirksam erhalten, wodurch die Sprachverständlichkeit hier gegenüber dem Mikrophonsignal merklich verbessert wird. Nach einer langen Sprachpause kommt es wegen der anschließenden Pausenunterdrückung nicht mehr zu einer Schwellenunterschreitung (Fig. 24). Aus diesem Grunde ist die in Fig. 25 unten am Ende der Szene festgehaltene Impulsantwort auf der Basis der gespeicherten Koeffizienten zeitlich relativ veraltet, weil sie bereits ca. 2,3 s vorher aktuell war (215 Blöcke  $\cdot 10,7$  ms). Wie der weist die aktuelle Impulsantwort (Fig. 25 oben) starke von der Sprach-eingabe herrührende Störungen auf. Wie ein Vergleich mit der ähnlichen Szene Z1 nach Figuren 11 bis 15 zeigt, ist trotz des stark verringerten Rechenaufwandes die Qualität der Störbeeinflussung unverändert hoch.

[0039] Die letzte Szene Z4 nach Fig. 26 wurde ohne Sprach-eingabe erstellt und soll abschließend nochmals die Musikunterdrückungseigenschaften des beschriebenen FLMS-Algorithmus demonstrieren. Nach ca. 18000 samples bzw. 1,5 s wird - wie aus Fig. 26 unten ersichtlich - die Musik wirksam unterdrückt. Diese Eigenschaft wird bis zum Szenenende mit unveränderter Qualität beibehalten. Fig. 27 zeigt auf, daß das

Sprachpausen-Größe  $t_{pa}$  überwiegend unter der Schwelle  $t_h$  bleibt. Die Zeiten, in welchen auf die geschilderten Koeffizienten zurückgegriffen wird, sind demnach nur sehr kurz. Impulsantwort und Übertragungsfunktion aus aktuellen Koeffizienten sind daher im wesentlichen mit den entsprechenden Verfahren aus Sprachpausen-Koeffizienten identisch.

#### Patentansprüche

1. Verfahren zur Störfreilegung eines Mikrophonsignals von Anteilen eines Quellsignals, daß als Referenzsignal ( $x$ ) vorliegt und nach Durchlaufen einer Übertragungsfunktion mit a priori unbekannter Übertragungsfunktion ( $G$ ) sich im Mikrophonsignal als Störsignal ( $y$ ) einem Sprachsignal ( $s$ ) überlagert, durch adaptive Nachbildung des Störsignals und Kompensation des tatsächlichen und des nachgebildeten Störsignals in einem Ausgangssignal, wobei das Mikrophonsignal geteilt ist in den Frequenzbereich transformiert, die Kompensation an Signalen im Frequenzbereich vorgenommen und das im Frequenzbereich vorliegende Ausgangssignal zur Adaption der Nachbildung mit dem im Frequenzbereich vorliegenden Referenzsignal verknüpft wird.
2. Verfahren nach Anspruch 1, dadurch gekennzeichnet, daß das Ausgangssignalspektrum in den Zeitbereich transformiert, das Zeitsignal durch Voranstellen von Nullwerten auf doppelte Länge gebracht, in den Frequenzbereich rücktransformiert und der Nachbildung der Übertragungsfunktion zugrundegelegt wird.
3. Verfahren nach Anspruch 1, dadurch gekennzeichnet, daß das Ausgangssignalspektrum mit dem Spektrum eines Hamming-Zeitfensters geteilt wird und der Nachbildung der Übertragungsfunktion zugrundegelegt wird.
4. Verfahren nach einem der Ansprüche 1 bis 3, dadurch gekennzeichnet, daß zur Nachbildung des Störsignals eine adaptive Filterfunktion eines Nachbildungstellers auf das Referenzsignal angewandt wird.
5. Verfahren nach Anspruch 4, dadurch gekennzeichnet, daß die Filterfunktion durch einen Koeffizientenvektor vorgegeben wird, dessen Koeffizienten adaptiv eingestellt werden.
6. Verfahren nach Anspruch 4 oder 5, dadurch gekennzeichnet, daß das Auftreten eines Sprachsignals im Mikrophonsignal detektiert wird und bei Auftreten eines Sprachsignals die vor Auftreten des Sprachsignals eingestellte Filterfunktion zur Bildung des Ausgangssignals beibehalten wird.
7. Verfahren nach Anspruch 6, dadurch gekennzeichnet, daß auch bei Detektion eines Sprachsignals die adaptive Nachbildung einer aktuellen Filterfunktion zusätzlich zur Bildung des Ausgangssignals fortgeführt wird.
8. Verfahren nach Anspruch 7, dadurch gekennzeichnet, daß das Auftreten eines Sprachsignals aus einer Veränderung der aktuellen Filterfunktion detektiert wird.
9. Verfahren nach Anspruch 8, dadurch gekennzeichnet, daß die Veränderung der aktuellen Filterfunktion für die Detektion des Auftretens eines Sprachsignals zeitlich gegliedert wird.
10. Verfahren nach einem der Ansprüche 4 bis 9, dadurch gekennzeichnet, daß die Filterfunktion in mehrere Teilfilterfunktion zu aufeinanderfolgenden Abschnitten einer Gesamt-Impulsantwort aller Teilfilter aufgespalten ist und auf Referenzsignalspektren zu zeitlich versetzten Zeitsegmenten des segmentierten Referenz-Zeitsignals angewandt wird.
11. Verfahren nach Anspruch 10, dadurch gekennzeichnet, daß die Adaption der Filterfunktion für die Teilfilter parallel durchgeführt wird.
12. Verfahren nach Anspruch 10, dadurch gekennzeichnet, daß die Adaption der Filterfunktion für die einzelnen Teilfilter zeitsequentiell durchgeführt wird.

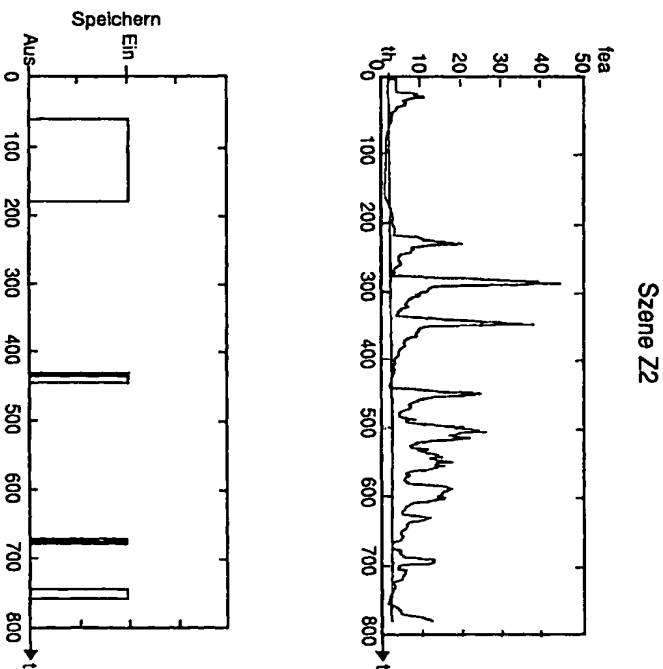


FIG.21



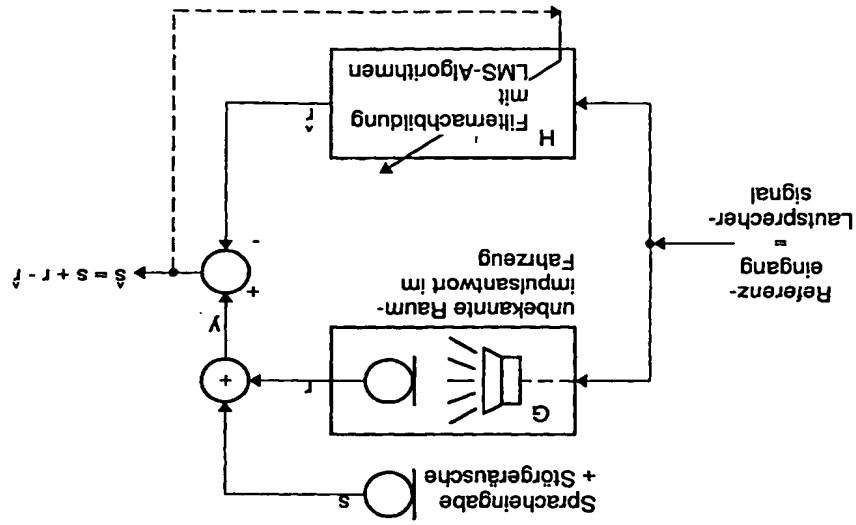
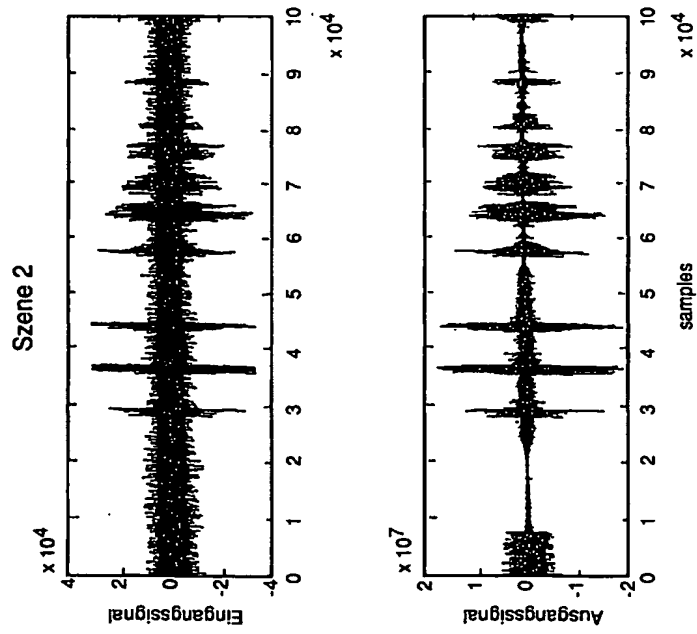
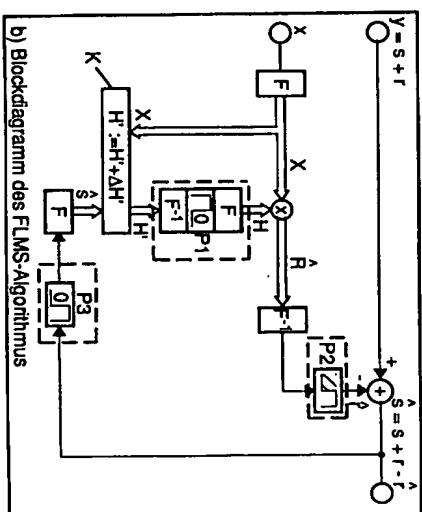
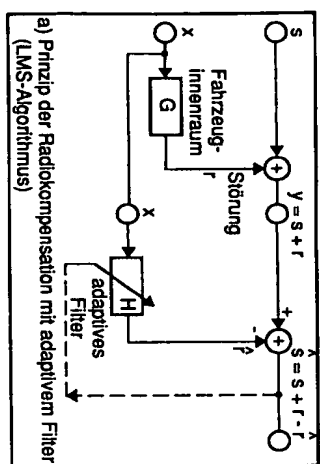
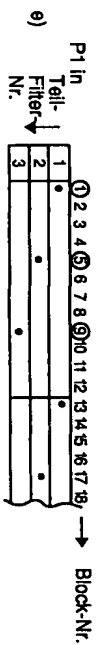
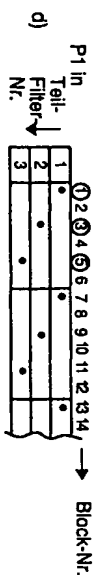
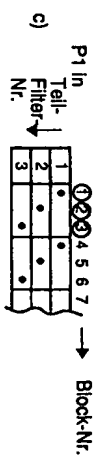
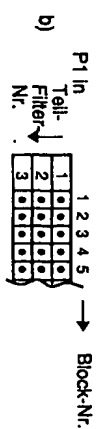
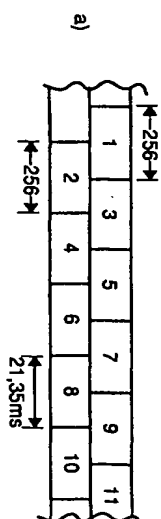


FIG.20

FIG.1



**FIG.2**



- Numerierung der überlappenden FFT-Blöcke
- Projektion P1 parallel in 3 Teilfiltern
- Projektion P1 sequentiell in 3 Teilfiltern (Blockmodule 3)
- Projektion P1 sequentiell in 3 Teilfiltern (Blockmodule 6)
- Projektion P1 sequentiell in 3 Teilfiltern (Blockmodule 12)

**FIG.19**

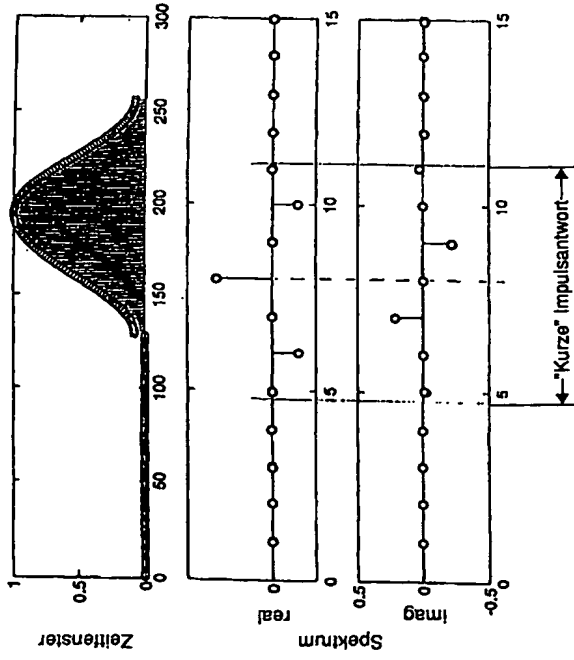


FIG.18

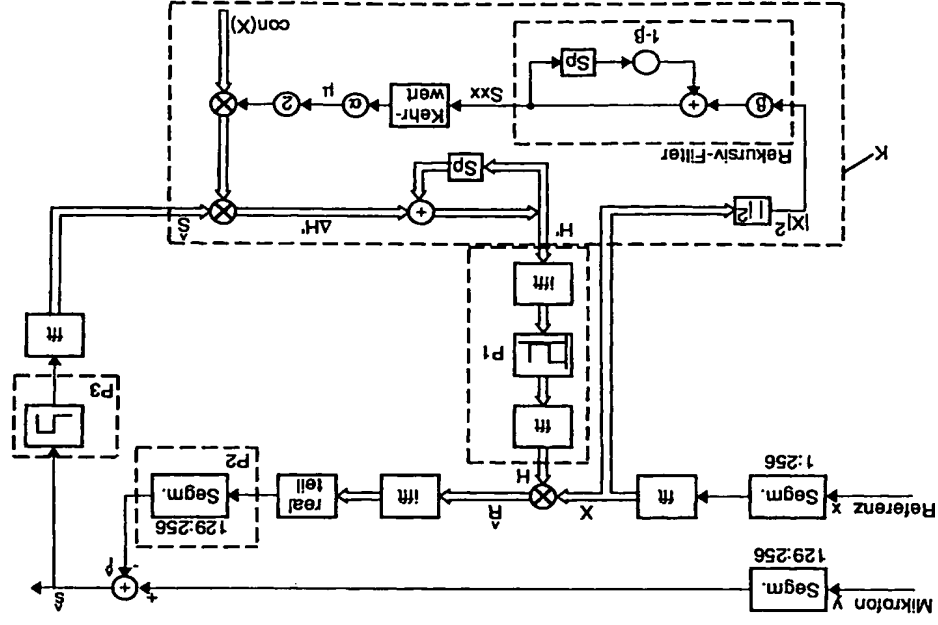


FIG.3

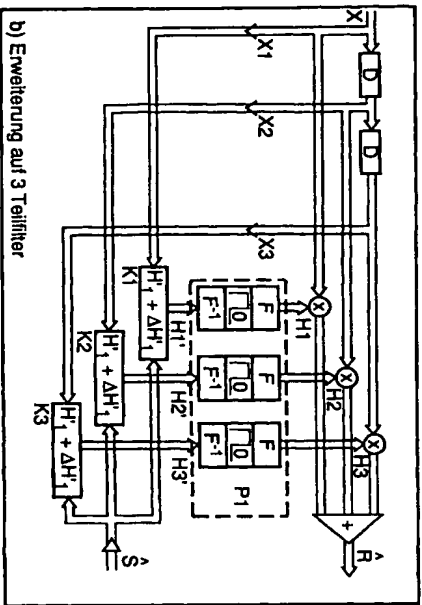
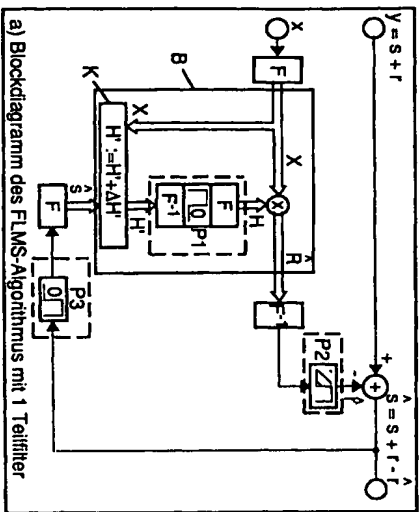


FIG.4

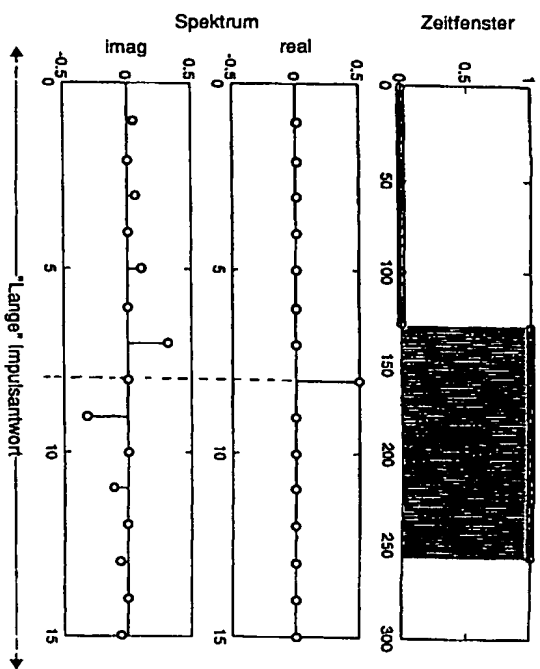


FIG.17

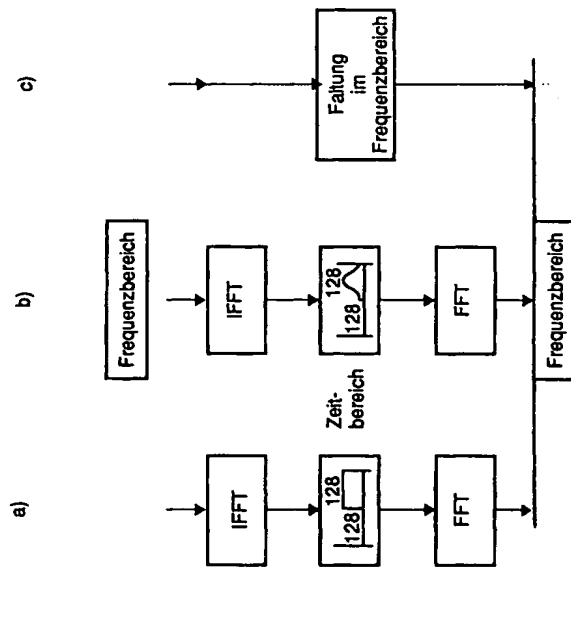


FIG.16

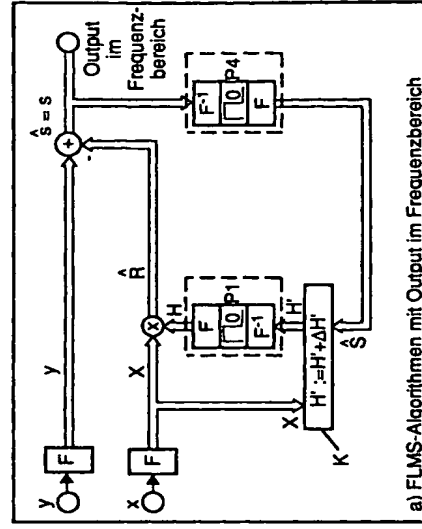
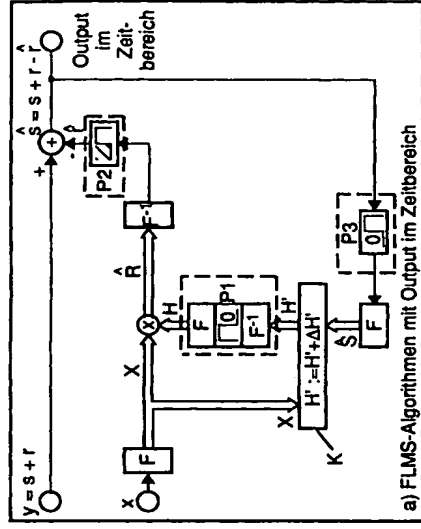
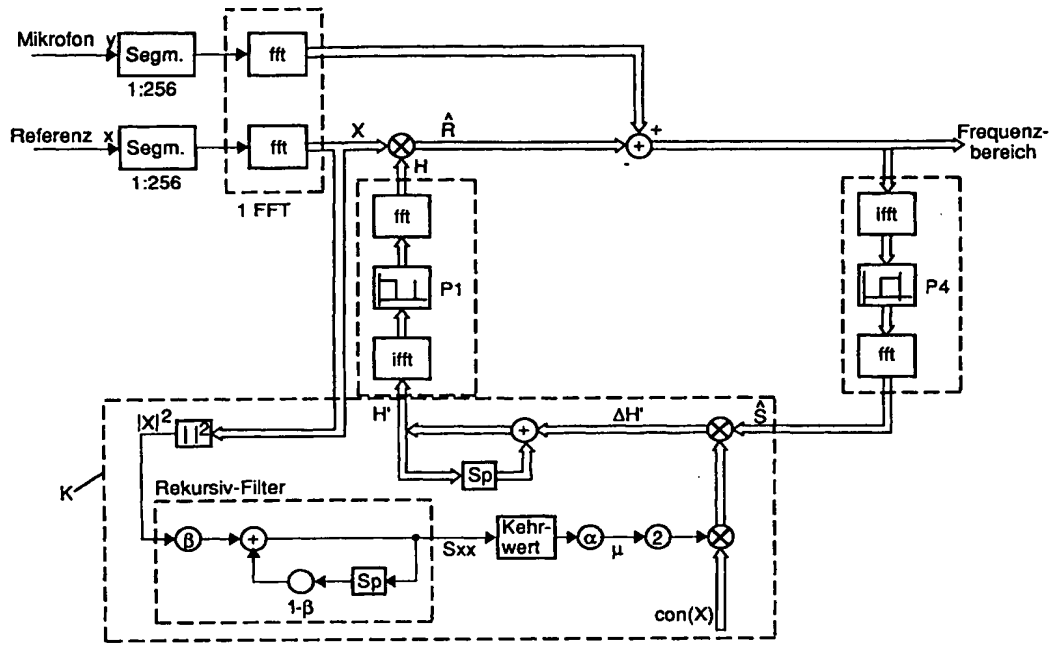
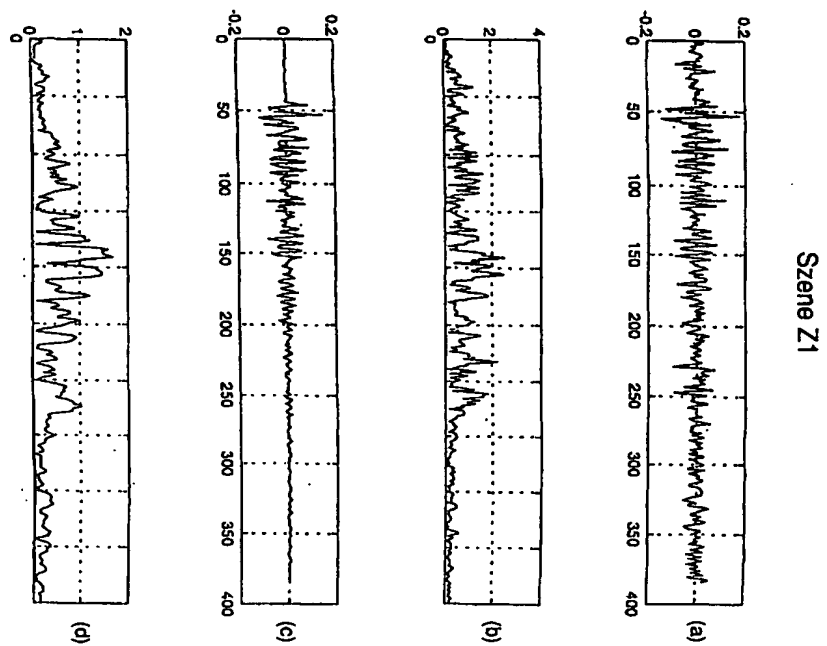


FIG.5



**FIG. 6**



**FIG. 15**



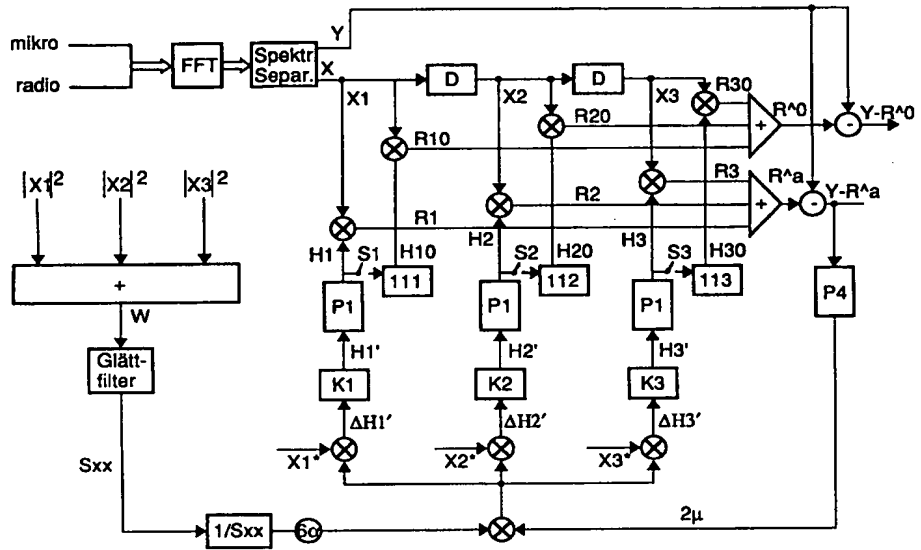


FIG. 8

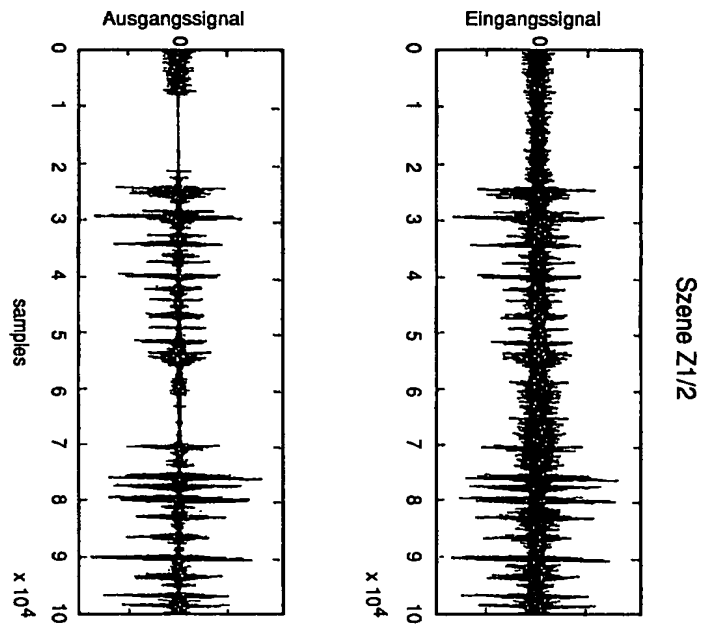


FIG. 13



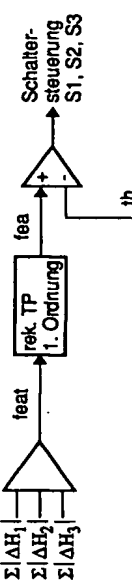
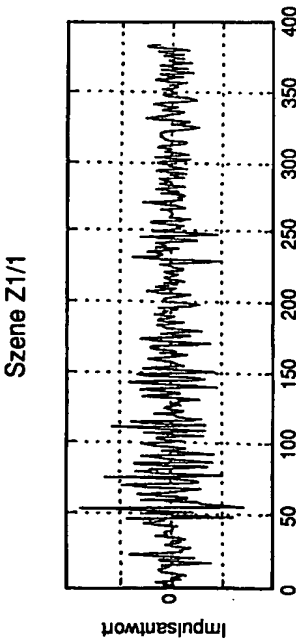


FIG.8a

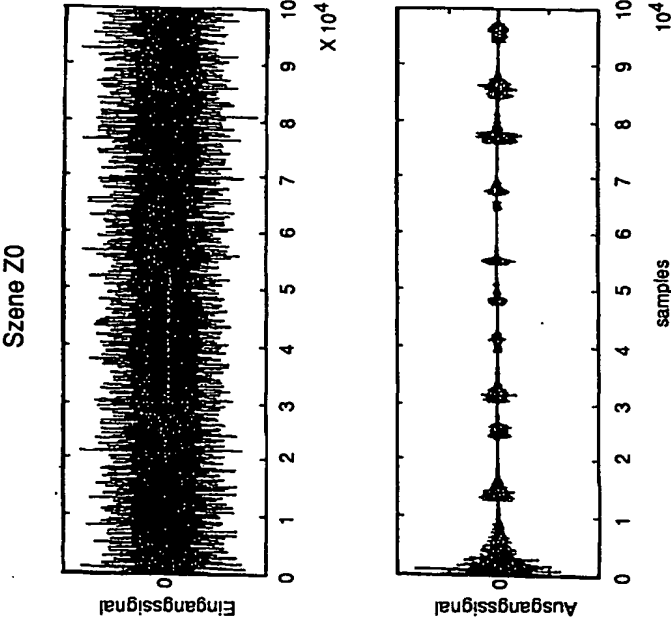
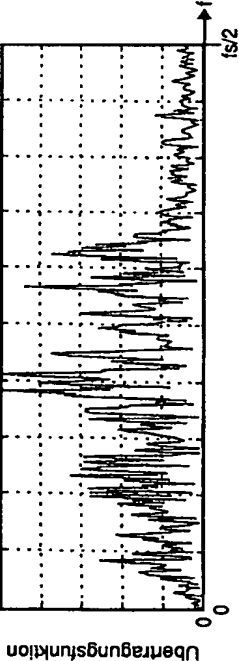
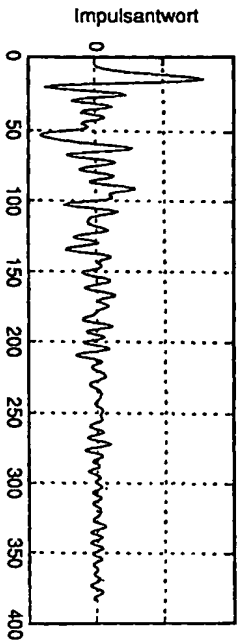


FIG.9

FIG.12



Szene Z0



Szene Z1/1

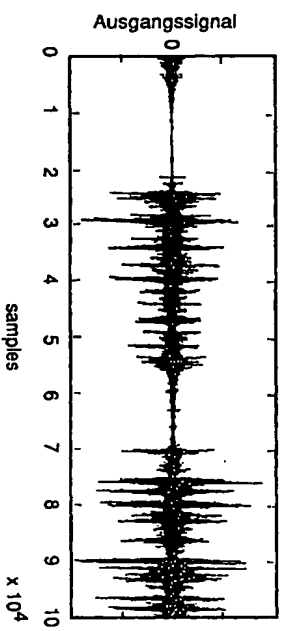
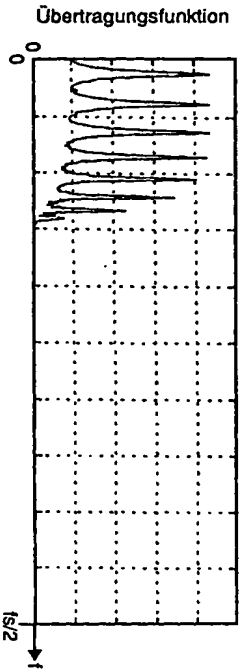
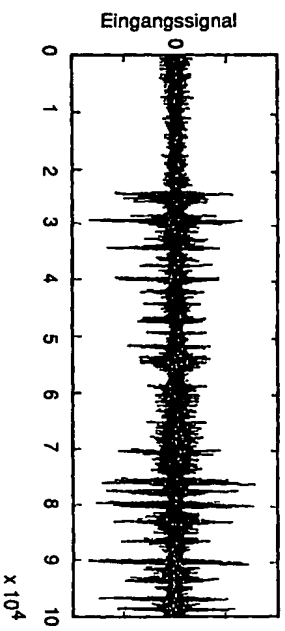


FIG.10

FIG.11

This Page is inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☒ BLACK BORDERS
- ☒ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☒ FADED TEXT OR DRAWING
- ☐ BLURED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☒ COLORED OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REPERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning documents *will not* correct images  
problems checked, please do not report the  
problems to the IFW Image Problem Mailbox**

**This Page Blank (uspto)**